3-Ч

УДК 004.89:004.93

Г.Г. Четвериков, И.Д. Вечирская, Т.Н. Фёдорова
Харьковский национальный университет радиоэлектроники, г. Харьков
Украина, 61166, г. Харьков, пр. Ленина, 14, tanja_fyodorova@mail.ru

Математическое описание словоизменительной классификации грамматических категорий в итальянском языке

G.G. Chetverikov, I.D. Vechirska, T.M. Fedorova
Kharkov National University of Radioelectronics, c. Kharkov
Ukraine, 61166, Kharkov, Lenin Avenue, 14, tanja_fyodorova@mail.ru

Mathematical Description of Inflectional Classification of Grammatical Categories in Italian

Г.Г. Четвериков, І.Д. Вечірська, Т.М. Федорова
Харківський національний університет радіоелектроніки, м. Харків
Україна, 61166, м. Харків, пр. Леніна, 14, tanja_fyodorova@mail.ru

Математичний опис словозмінної класифікації граматичних категорій в італійській мові

В статье рассматривается проблема построения формализованных словоизменительных классификаций на примере итальянского языка. Приведена система признаков (грамматических категорий), которые определяют словоизменение указанных классов слов. При разработке модели учитываются все значения, которые эти признаки (категории) могут принимать.
Ключевые слова: словоизменение, грамматические категории, признаки, структура.

The article considers the problem of constructing a formal inflectional classification as an example of Italian language. Shows the system of attributes (grammatical categories), which determine the inflection of the word classes. In developing the model takes into account all the values that these attributes (categories) can take.
Key words: inflection, grammatical categories, attributes, structure.

У статті розглядається проблема побудови формалізованих словозмінних класифікацій на прикладі італійської мови. Наведено систему ознак (граматичних категорій), які визначають словозміну зазначених класів слів. При розробці моделі враховуються всі значення, які ці ознаки (категорії) можуть приймати.
Ключові слова: словозміна, граматичні категорії, ознаки, структура.

Введение

В последнее время наряду с традиционными лексико-грамматическими исследованиями в лингвистике все большую актуальность приобретает вопрос построения формализованных словоизменительных классификаций. Это связано как с теоретическими задачами грамматики по разноплановому и глубокому описанию языковых систем, так и с необходимостью построения эффективных лингвистических технологий. Основным требованием к описанию языковой системе, ориентированной на применение в лингвистических технологиях, является максимальная формализация этого

описания, доведение результатов соответствующих исследований до уровня формальных моделей и лингвистических алгоритмов, адаптированных к использованию в современных языково-информационных системах.

Актуальной является проблема создания морфологических классификаций слов и на этой основе – средств грамматической идентификации и маркировки лингвистических структур в естественно-языковых текстах. В этой связи необходимыми оказываются компьютерные программы автоматической грамматической идентификации лексических единиц, способные самостоятельно, без участия человека, приписывать параметры грамматического значения текстовым словоформам. И без формализованных словоизменительных классификаций создания таких средств невозможно. [1-6].

Целью данной работы является разработка математической модели словоизменительной классификации частей речи итальянского языка для электронного словаря.

Моделирование словоизменения

Романские языки принадлежат к флективным языкам с сильной тенденцией к аналитизму. Морфологическое выражение нерегулярно (имеются случаи невыраженности грамматических категорий и морфологических омонимов). Существительное имеет категории числа, рода (мужской и женский род; латинский средний перераспределился между ними). Особенностью романских языков является разнообразие форм артикля. В местоимениях сохраняются элементы системы падежей. Отличительная черта романских языков – наличие двух рядов объектных местоимений: самостоятельных и служебных. Прилагательные согласуются с существительным в роде и в числе во всех позициях, но некоторые не изменяются в роде. Романские языки характеризуются разветвленной системой глагольных форм. Широкое развитие получили аналитические формы, состоящие из вспомогательного глагола и безличных форм (причастия, инфинитива, герундия). Имеются активный и пассивный залоги, а также местоименная форма, выражающая возвратное (и косвенно-возвратное), взаимное (и косвенно-взаимное), пассивное или неопределенно-личное значения. Своеобразны в романских языках неличные формы глагола (инфинитив, герундий, причастие II, в некоторых языках и причастие I). Порядок слов в ряде случаев фиксирован: в сложной глагольной форме вспомогательный глагол предшествует причастию (инфинитиву). Прилагательное обычно следует за существительным, тогда как детерминативы предшествуют имени.

Итальянский язык – один из романских языков (итало-романская подгруппа), в котором явление словоизменения является характерным для основных значимых частей речи. Определенное количество словоизменительных форм образуется синтетически, т.е. путем присоединения к основе слова флексий, которые содержат в себе информацию о конкретных значения одного или нескольких грамматических категорий. Также существует часть грамматических форм, которые образуются аналитически (грамматические формы состоят более чем из одного слова).

Приведем систему признаков (грамматических категорий) для итальянского языка, которые определяют словоизменение указанных классов слов и значения, которые эти признаки (категории) могут принимать.

Имя существительное (Il sostantivo) (parola (слово), domando (вопрос))

• Род (gènere)– мужской (maschile), женский (femminile);

• Число (nùmero) – единственное (singolare), множественное (plurale).

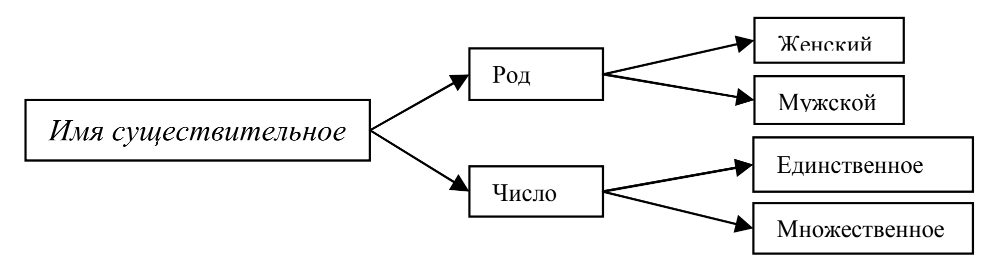


Рисунок 1 – Структура грамматических категорий имени существительного

Артикль (L' articolo) (il, l', lo, la, i, gli, le, un, una, uno, un')
• Вид* – определенный (determinato), неопределенный (indeterminato);
• Род (gènere)– мужской (maschile), женский (femminile);
• Число (nùmero) – единственное (singolare), множественное (plurale).

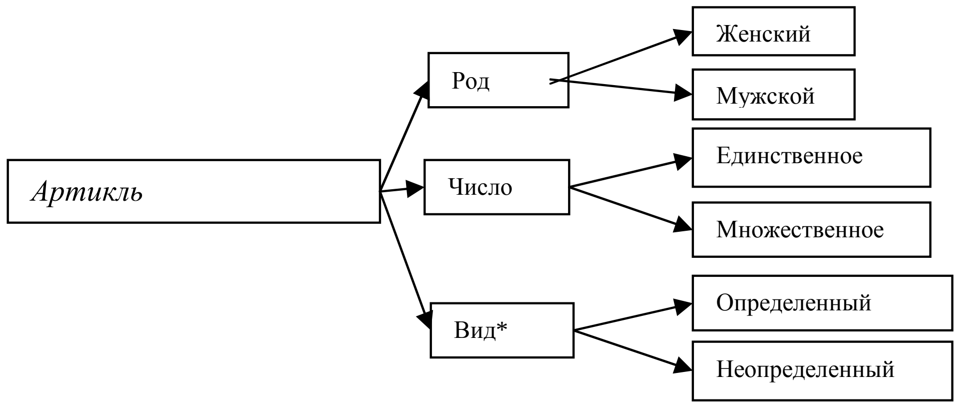


Рисунок 2 – Структура грамматических категорий артикля

Имя прилагательное (L' aggettivo) (verde (зеленый), amichevole (дружеский), mio (мой), tuo (твой), questo (этот), quello (тот), alcuno (несколько), diversi (многие), quanto (сколько), quale (какой, который из))

• Тип* – качественные (qualitativi), притяжательные (possessivi), указательные (dimostrativi), неопределенные (indefiniti), вопросительные (interrogativi), восклицательные (esclamativo);

• Род (gènere)– мужской (maschile), женский (femminile);

• Число (nùmero) – единственное (singolare), множественное (plurale).

•

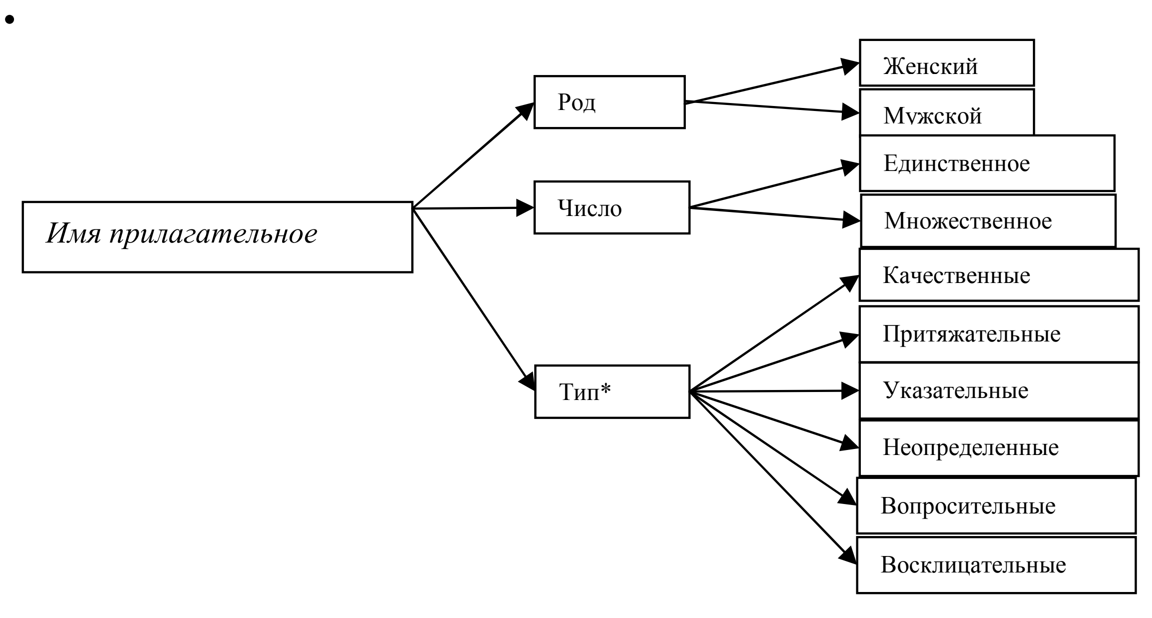


Рисунок 3 – Структура грамматических категорий имени прилагательного

Имя числительное (Il numerale) (cinque (пять), ventidue (двадцать два), primo (первый), ventiduesimo (двадцать второй), una decina (десяток), una dozzina (дюжина))

• Тип* – количественные (quantitative), порядковые (ordinali), существительные неточного количества (sostantivi inesatta numero);

• Род (gènere)– мужской (maschile), женский (femminile);

• Число (nùmero) – единственное (singolare), множественное (plurale).

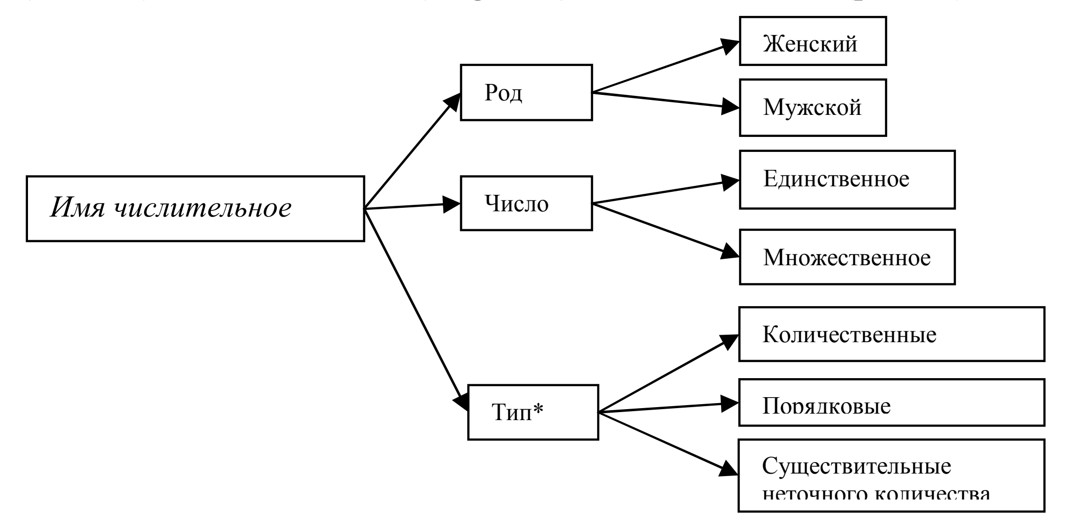


Рисунок 4 – Структура грамматических категорий имени числительного

Местоимение (Il pronome) (lei (она), loro (они), questo (это), quello (что), chiunque (любой), tutto (всё, весь), che (который), chi (кто), che cosa (что), quale (какой))

• Тип* – личные (personali), указательные (dimostrativi), неопределенные (indefiniti), относительные (relativi), вопросительные (interrogativi);

• Род (gènere)– мужской (maschile), женский (femminile);

• Число (nùmero) – единственное (singolare), множественное (plurale);

• Лицо (persona) – 1 (prima), 2 (secondo), 3 (terzo).

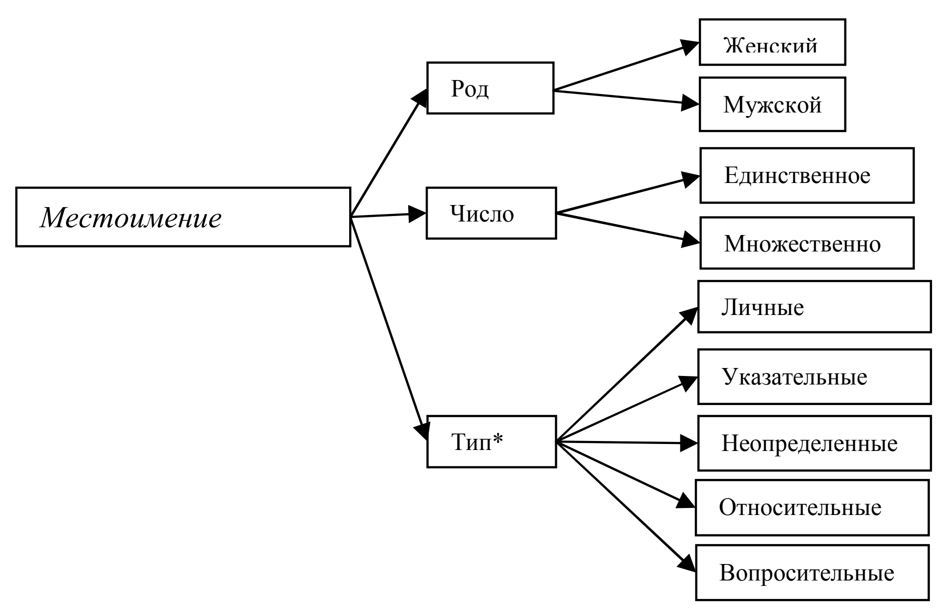


Рисунок 5 – Структура грамматических категорий местоимения

Глагол (Il verbo) (essere (быть), avere (иметь), telefonare (звонить), prendere (брать), dormire (спать), finire (заканчивать))

• Лицо (persona) – 1 (prima), 2 (secondo), 3 (terzo);

• Число (nùmero) – единственное (singolare), множественное (plurale);

• Наклонение (modo) – изъявительное (indicativo), сослагательное (congiuntivo), условное (condizionale), повелительное (imperativo)

• Время (tempo) – настоящее (presente), прошедшее (passato), будущее (futuro semplice), прошедшее незаконченное время (imperfetto), прошедшее законченное время (passato remoto), прошедшее ближайшее время (passato prossimo), предпрошедшее время (trapassato prossimo), предпрошедшее непосредственное время (trapassato remoto), сложное будущее время (futuro anteriore), будущее время прошедшего (condizionale passato).

• Залог (voce) – действительный (vera), страдательный (passiva).

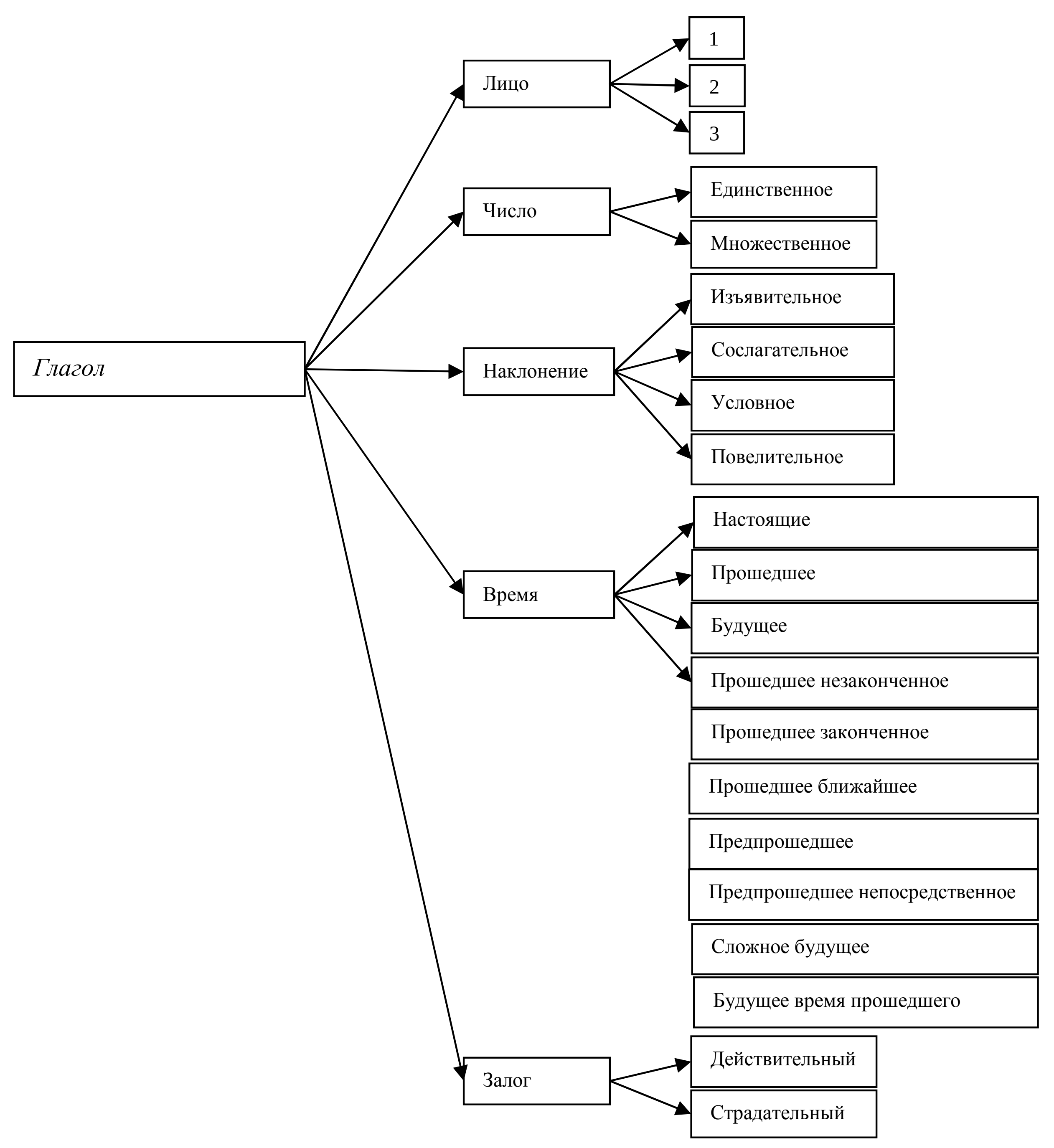


Рисунок 6 – Структура грамматических категорий глагола

По значениям своих классификационных категорий грамматические классы делятся на подклассы (в приведенной выше системе признаков для каждого из грамматических классов их классификационные признаки обозначены символом «*»). Количество подклассов зависит от того, сколько значений может принимать данная классификационная категория. Например, класс артиклей делится на два подкласса по признаку вида; класс прилагательных по типу делится на 6 подклассов; класс числительных по типу делится на 3 подкласса; класс местоимений по типу делится на 5 подклассов.

Пусть W – множество склоняемых слов итальянского языка. Разобьем множество W на подмножества (классы слов), которые взаимно не пересекаются:

$$W = \bigcup_{j=1}^{6} W_j \,, \tag{1.1}$$

где W_1 – имя существительное, W_2 – артикль, W_3 – имя прилагательное, W_4 – имя числительное, W_5 – местоимение, W_6 – глагол.

Омонимию будем считать снятой, а омонимы промаркированными, так что

$$W_{j1} \bigcap W_{j2} = \varnothing \text{ при } j_1 \neq j_2,\ j_1, j_2 = \overline{1,6}.$$

Некоторые классы слов по своим классификационным признакам подразделяются на подклассы. Подклассы будем обозначать W_j^k.

Артикли по значению грамматической категории *вид* делятся на два подкласса:

$$W_2 = W_2^1 \cup W_2^{\,2},$$

где W_2^1 – определенный артикль, $W_2^{\,2}$ – неопределенный артикль.

Подклассы W_j^k будем называть грамматическими классами артиклей и обозначать их в дальнейшем P_i, $i = 2,3$: $P_i \equiv W_j^{\,k}, \quad j = 2, k = 1,2$.

Имена прилагательные по значению грамматической категории *тип* делятся на следующие подклассы:

$$W_3 = W_3^1 \cup W_3^{\,2} \cup W_3^{\,3} \cup W_3^{\,4} \cup W_3^{\,5} \cup W_3^{\,6},$$

где W_3^1 – качественные прилагательные, $W_3^{\,2}$ – притяжательные прилагательные, W_3^3 – указательные прилагательные, $W_3^{\,4}$ – неопределенные прилагательные, W_3^5 – вопросительные прилагательные, $W_3^{\,6}$ – восклицательные прилагательные.

Подклассы W_3^k (k=1,2,...,6) будем называть грамматическими классами прилагательных. Обозначим $P_i \equiv W_3^{\,k}, \quad i = \overline{4,9}; k = \overline{1,6}$.

Имена числительные по значению грамматической категории *тип* делятся на следующие подклассы:

$$W_4 = W_4^1 \cup W_4^{\,2} \cup W_4^{\,3},$$

где W_4^1 – количественные числительные, $W_4^{\,2}$ – порядковые числительные, $W_4^{\,3}$ – существительные неточного количества.

Подклассы W_4^k (k=1,2,3) будем называть грамматическими классами числительных. Обозначим $P_i \equiv W_4^{\,k}, \quad i = 10,11,12; k = 1,2,3$.

Местоимения по значению грамматической категории *тип* делятся на следующие подклассы:

$W_5 = W_5^1 \cup W_5^{\,2} \cup W_5^{\,3} \cup W_5^{\,4} \cup W_5^{\,5}$, где W_5^1 – личные местоимения, $W_5^{\,2}$ – указательные местоимения, W_5^3 – неопределенные местоимения, $W_5^{\,4}$ – относительные местоимения, W_5^5 – вопросительные местоимения.

Подклассы W_5^k (k=1,2,...,5) будем называть грамматическими классами местоимений. Обозначим $P_i \equiv W_5^{\,k}, \quad i = 13,14,...,17; k = 1,2,...,5$.

В результате для каждого $x \in W$ определяется принадлежность к определенному грамматическому классу $P_i, i = 1,2,...18$.

Всего имеем 18 грамматических классов:

$$W = \bigcup_{i=1}^{18} P_i \,,\; P_{i_1} \cap P_{i_1} = \varnothing, i_1 \neq i_2, i_1, i_2 = \overline{1,18}.$$

Внутри грамматических классов выделяем парадигматические классы. Это означает, что произвольная лексема *x* представляется в виде комбинации неизменяемой и изменяемой составляющих (квазиосновы и квазифлексии): *x=c(x)*f(x),* где *c(x)* – квазиоснова, *f(x)* – квазифлексия, ***** – конкатенация [7].

Квазифлексия помимо только флексии может включать часть основания, в которой происходит чередование. Например, итальянский глагол '*pescare*' – ловить рыбу.

io//я	pes<u>c</u>o	noi//мы	pes<u>ch</u>iamo
tu//ты	pes<u>ch</u>i	voi//вы	pes<u>c</u>ate
lui/lei//он/она	pes<u>c</u>a	loro//они	pes<u>c</u>ano

Таким образом, моделирование словоизменений происходит по следующему правилу: для всех слов, принадлежащих к одному парадигматическому классу, парадигматические формы в конкретных грамматических значениях строятся путем присоединения к квазиоснове слова квазифлексий, соответствующие этим грамматическим значениям.

Выводы

В статье был проведен анализ грамматики итальянского языка. Были определены парадигматические типы и словоизменительные параметры итальянского языка. Лексический состав языка был разбит на грамматические классы. Было установлено соответствие между парадигматическими типами, грамматическими классами и словоизменительными категориями. Внутри грамматических классов были выделены парадигматические классы. Таким образом, была сформирована база для построения модели грамматической лексикографической системы итальянского языка. На основе полученных результатов была построена математическая модель словоизменительной классификации частей речи итальянского языка

Разработанная модель словоизменительной классификации итальянской лексики является удобной для использования ее в электронных словарях и соответствует всем требованиям построения подобных словарей.

Литература

1. Широков К.В. Іменна словозміна у сучасній турецькій мові / К.В. Широков. – К. : Довіра, 2009. – 318 с.
2. Бондаренко М.Ф. Основи території синтезу надшвидкодіючих структур мовних систем штучного інтелекту / М.Ф. Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков. – К., 1997.-263 с.
3. Бондаренко М.Ф. Основи території багатозначних структур і кодування в системах штучного інтелекту / М.Ф. Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков. – Х. : Фактор-Друк, 2003. – 336 с.
4. Четвериков Г.Г. Алгебро-логические средства моделирования естественного языка. Сообщение 3 / Г. Г. Четвериков // Бионика интеллекта. – 2007. – № 1 (66) – С. 48-52.

5. Четвериков Г.Г. Малий тлумачний словник з інтелектуальних систем : навч. посібник / Г.Г. Четвериков. – К. : ІЗМН, 1997. – 64 с.
6. Четвериков Г.Г. Малий тлумачний словник з інтелектуальних систем (Доповнення) : навч. посібник / Г.Г. Четвериков. – Х. : ХТУРЕ, 1999. – 56 с.
7. Широков В.А. Елементи лексикографії / В.А. Широков. – К. : Довіра, 2005.-303 с.

Literatura

1. Shyrokov K.V. Imenna slovozmina u suchasnij turec'kij movi. K.: Dovira. 2009. 318 s.
2. Bondarenko M.F. Osnovy terytorii syntezu nadshvydkodijuchyh struktur movnyh system shtuchnogo intelektu. K.. 1997. 263s.
3. Bondarenko M. F. Osnovy teritorii bagatoznachnyh struktur i koduvannja v systemah shtuchnogo intelektu. H.: Faktor-Druk. 2003. 336s.
4. Chetverikov G.G. Bionika intellekta. 2007. № 1 (66). S. 48-52.
5. Chetverikov G.G. Malyj tlumachnij slovnyk z intelektual'nyh system: Navch. Posibnyk. K.: IZMN. 1997. 64 s.
6. Chetverikov G.G. Malyj tlumachnij slovnyk z intelektual'nyh system (Dopovnennja): Navch. Posibnik. H.: HTURE. 1999. 56 s.
7. Shyrokov V.A. Elementy leksykografii. K.: Dovira. 2005. 303 s.

RESUME

G.G. Chetverikov, I.D. Vechirska, T.M. Fedorova

Mathematical Description of Inflectional Classification of Grammatical Categories in Italian

In the article, Italian grammar is analyzed. Paradigmatic types and inflectional parameters are identified. The lexical structure of language is divided into grammar classes. It is established the correlation between the paradigmatic types, grammar classes and inflectional categories. Within the grammatical classes, paradigmatic classes are identified. Thus, the basis for constructing the model of the grammatical lexicographic system of Italian has been made. On the basis of the results, mathematical model for the inflectional classification of Italian parts of speech is constructed.

The developed model of inflectional classification of the Italian language is suitable for use in electronic dictionaries and corresponds to the all requirements for constructing such dictionaries.

Статья поступила в редакцию 03.06.2012.